CLAIMS

What is claimed is:

1. A method of identifying a set of informative genes whose expression correlates with a class distinction between samples, comprising the steps of:

a) sorting genes by degree to which their expression in said samples correlate with a class distinction; and

b) determining whether said correlation is stronger than expected by chance;

wherein a gene whose expression correlates with a class distinction more strongly than expected by chance is an informative gene, thereby identifying a set of informative genes.

2. The method of Claim 1, wherein the class is a known class.

3. The method of Claim 2, wherein the class distinction is a disease class distinction.

4. The method of Claim 3, wherein the disease class distinction is a cancer class distinction.

5. The method of Claim 4, wherein the cancer class distinction is selected from the group consisting of a leukemia class distinction, a brain tumor class distinction and a lymphoma class distinction.

6. The method of Claim 1, wherein step (a) is carried out by neighborhood analysis.

7. The method of Claim 6, wherein said neighborhood analysis comprises the steps of:

a) defining an idealized expression pattern corresponding to a gene, wherein said idealized expression pattern is expression of said gene that is uniformly high in a first class and uniformly low in a second class; and

b) determining whether there is a high density of genes having an expression pattern similar to said idealized expression pattern, as compared to an equivalent random expression pattern, wherein the high density of genes are genes having a high statistical significance in a permutation test.

8. The method of Claim 7, wherein the signal to noise routine is:

$$P(g,c)= (\mu_1(g)-\mu_2(g))/(\sigma_1(g) + \sigma_2(g)),$$

wherein g is the gene expression value; c is the class distinction, $\mu_1(g)$ is the mean of the expression levels for g for the first class; $\mu_2(g)$ is the mean of the expression levels for g for the second class; $\sigma_1(g)$ is the standard deviation for the first class; and $\sigma_2(g)$ is the standard deviation for the second class.

9. A method of assigning a sample to a known or putative class, comprising the steps of:

a) determining a weighted vote for one of the classes of one or more informative genes in said sample in accordance with a model built with a weighted voting scheme, wherein the magnitude of each vote depends on the expression level of the gene in said sample and on the

degree of correlation of the gene's expression with class distinction; and

b) summing the votes to determine the winning class.

10. The method of Claim 9, wherein the weighted voting scheme:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a vote for the second class.

11. The method of Claim 10, wherein a set of informative genes whose expression correlates with a class distinction between samples is identified.

12. The method of Claim 11, wherein identifying a set of informative genes, comprises the steps of:

a) sorting genes by degree to which their expression in said samples correlate with a class distinction; and

b) determining whether said correlation is stronger than expected by chance;

wherein a gene whose expression correlates with a class distinction more strongly than expected by chance is an informative gene, thereby identifying a set of informative genes.

13. The method of Claim 12, wherein the set of informative genes is determined with a signal to noise routine is:

$$P(g,c)= (\mu_1(g)-\mu_2(g))/(\sigma_1(g) + \sigma_2(g)),$$

wherein g is the gene expression value; c is the class distinction, $\mu_1(g)$ is the mean of the expression levels for g for a first class; $\mu_2(g)$ is the mean of the expression levels for g for a second class; $\sigma_1(g)$ is the standard deviation for the first class; and $\sigma_2(g)$ is the standard deviation for the second class.

14. A method of assigning a sample to a known or putative class, comprising the steps of:

a) determining a weighted vote for one of the classes for one or more informative genes in said sample in accordance with a model built with a weighted voting scheme, wherein the magnitude of each vote depends on the expression level of the gene in said sample and on the degree of correlation of the gene's expression with class distinction; and

b) summing the votes to determine the winning class and a prediction strength,

wherein said sample is assigned to the winning class if the prediction strength is greater than a prediction strength threshold.

15. The method of Claim 14, wherein the prediction strength is determined by:

$$(V_{win} - V_{lose}) / (V_{win} + V_{lose}),$$

wherein V_{win} and V_{lose} are the vote totals for the winning and losing classes, respectively.

16. The method of Claim 15, wherein the number of informative genes used in the weighted voting scheme is at least 50.

17. The method of Claim 16, wherein the known class is a known disease class.

18. The method of Claim 17, wherein the disease class is a cancer disease class.

19. The method of Claim 18, wherein the cancer disease class is Acute Lymphoblastic Leukemia (ALL) or Acute Myeloid Leukemia (AML).

20. The method of Claim 19, wherein the informative genes is selected from a group consisting of: C-myb, Proteasome iota, MB-1, Cyclin, Myosin light chain, Rb Ap48, SNF2, HkrT-1, E2A, Inducible protein, Dynein light chain, Topoisomerase II β, IRF2, TFIIEβ, Acyl-Coenzyme A, dehydrogenase, SNF2, ATPase, SRP9, MCM3, Deoxyhypusine synthase, Op 18, Rabaptin-5, Heterochromatin protein p25, IL-7 receptor, Adenosine deaminase, Fumarylacetoacetate, Zyxin, LTC4 synthase, LYN, HoxA9, CD33, Adipsin, Leptin receptor, Cystatin C, Proteoglycan 1, IL-8 precursor, Azurocidin, p62, CyP3, MCL1, ATPase, IL-8, Cathepsin D, Lectin, MAD-3, CD11c, Ebp72, Lysozyme, Properdin and Catalase.

21. The method of Claim 14, wherein the known class is a class of individuals who respond well to chemotherapy or a class of individuals who do not response well to chemotherapy.

22. A method of determining a weighted vote for an informative gene to be used in classifying a sample to be tested, comprising:

a) determining a weighted vote for one of the classes for one or more informative genes in said sample, wherein the magnitude of each vote depends on the expression level of the gene in said sample and

on the degree of correlation of the gene's expression with class distinction; and

b) summing the votes to determine the winning class.

23. The method of Claim 22, wherein the weighted vote determined according to:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a vote for the second class.

24. The method of Claim 23, wherein the vote for the first class is determined by obtaining a sum of the absolute values of the positive votes for the first class, and the vote for the second class is determined by obtaining a sum of the absolute values of the negative votes for the second class.

25. The method of Claim 24, wherein the weighted vote determined a portion of genes that are relevant for determining the classes.

26. The method of Claim 25, wherein a signal to noise routine, a Pearson correlation routine, or a Euclidean distance routine determines the relevant genes.

27. The method of Claim 26, wherein the signal to noise routine is:

$$P(g,c)= (\mu_1(g)-\mu_2(g))/(\sigma_1(g) + \sigma_2(g)),$$

wherein g is the gene expression value; c is the class distinction, $\mu_1(g)$ is the mean of the expression levels for g for the first class; $\mu_2(g)$ is the mean of the expression levels for g for the second class; $\sigma_1(g)$ is the standard deviation for the first class; and $\sigma_2(g)$ is the standard deviation for the second class.

28. A method for ascertaining a plurality of classifications from two or more samples, comprising:

a) clustering samples by gene expression values to produce putative classes; and

b) determining whether said putative classes are valid by carrying out class prediction based on putative classes and assessing whether new samples have a high prediction strength.

29. The method of Claim 29, wherein the clustering of the samples is performed according to a self organizing map.

30. The method of Claim 29, wherein the self organizing map is formed of a plurality of Nodes, N, and clusters the vectors according to a competitive learning routine.

31. The method of Claim 30, wherein the competitive learning routine is:

$$f_{i+1}(N) = f_i(N) + \tau(d(N,N_p), i)\,(P - f_i(N))$$

wherein i = number of iterations, N= the node of the self organizing map, τ = learning rate, P = the subject working vector, d = distance, N_p = node that is mapped nearest to P, and $f_i(N)$ is the position of N at i.

32. The method of Claim 28, wherein determining whether said putative classes are valid comprises:

a) determining a weighted vote for one of the classes for one or more informative genes in said sample, wherein the magnitude of each vote depends on the expression level of the gene in said sample and on the degree of correlation of the gene's expression with class distinction; and

b) summing the votes to determine the winning class.

33. The method of Claim 32, wherein the routine for building a model with a weighted voting scheme is:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a vote for the second class.

34. A method for classifying a sample obtained from an individual into a class, comprising:

a) assessing the sample for a level of gene expression for at least one gene; and

b) using a model built with a weighted voting scheme, classifying the sample as a function of relative gene expression level of the sample with respect to that of the model.

35. The method of Claim 34, wherein assessing the level of gene expression comprises assessing the level of expression of a gene product.

36. The method of Claim 35, wherein the individual has a disease, and the sample is classified into a class of the disease.

37. The method of Claim 36, wherein the disease is cancer.

38. The method of Claim 37, wherein the cancer is leukemia.

39. The method of Claim 38, wherein the leukemia is AML or ALL.

40. A method for classifying a sample into a cancer disease class, wherein the sample is obtained from an individual and the level of gene expression for at least one gene is determined, comprising, using a model built with a weighted voting scheme, classifying the sample as a function of relative gene expression level of the sample with respect to that of the model, to thereby classify the sample into the cancer disease class.

41. The method of Claim 17, wherein the cancer disease class is a leukemia class.

42. The method of Claim 18, wherein the leukemia class is AML or ALL.

43. A method for classifying a sample obtained from an individual, comprising:

a) subjecting the sample to at least one condition;

b) obtaining a gene expression product for two or more genes;

c) assessing the gene expression product for the genes to thereby determine the levels of the gene expression product for the genes;

d) using a computer model built with a weighted voting scheme, classifying the sample including comparing the gene expression levels of the sample to gene expression level of the model.

44. The method of Claim 43, wherein the genes assessed are the genes used to build the model.

45. In a computer system, a method for classifying at least one sample to be tested that is obtained from an individual, wherein gene expression values are determined for the sample to be tested, comprising:

a) receiving the gene expression values for the sample to be tested;

b) using a model built with a weighted voting scheme, classifying the sample including comparing the gene expression values of the sample to that of the model, to thereby produce a classification of the sample; and

c) providing an output indication of the classification.

46. The method of Claim 45, wherein the model is built according to:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a negative vote for the class.

47. The method of Claim 46, wherein the vote for the first class is determined by obtaining a sum of the absolute values of the positive votes for the first class, and the vote for the second class is determined by obtaining a sum of the absolute values of the negative votes for the second class.

48. The method of Claim 47, wherein the weighted voting scheme builds the model using a portion of genes that are relevant for determining the classes.

49. The method of Claim 48, wherein a signal to noise routine, a Pearson correlation routine, or a Euclidean distance routine determines the relevant genes.

50. The method of Claim 49, wherein the signal to noise routine is:

$$P(g,c)= (\mu_1(g)-\mu_2(g))/(\sigma_1(g) + \sigma_2(g)),$$

wherein g is the gene expression value; c is the class distinction, $\mu_1(g)$ is the mean of the expression levels for g for the first class; $\mu_2(g)$ is the mean of the expression levels for g for the second class; $\sigma_1(g)$ is the standard deviation for the first class; and $\sigma_2(g)$ is the standard deviation for the second class.

51. In a computer system, a method for classifying at least one sample obtained from an individual, comprising:

a) providing a model built by a weighted voting scheme;

b) assessing the sample for the level of gene expression for at least one gene, to thereby obtain a gene expression value for each gene;

c) using the model built with a weighted voting scheme, classifying the sample comprising comparing the gene expression level of the sample to the model, to thereby obtain a classification; and

d) providing an output indication of the classification.

52. The method of Claim 51, wherein the model is built by a routine having:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a negative vote for the class.

53. The method of Claim 52, wherein the vote for the first class is determined by obtaining a sum of the absolute values of the positive votes for the first class, and the vote for the second class is determined by obtaining a sum of the absolute values of the negative votes for the second class.

54. The method of Claim 53, wherein the weighted voting scheme builds the model using a portion of genes that are relevant for determining the classes.

55. The method of Claim 54, wherein a signal to noise routine, a Pearson correlation routine, or a Euclidean distance routine is used to determine the relevant genes.

56. The method of Claim 55, wherein the signal to noise routine is:

$$P(g,c)= (\mu_1(g)-\mu_2(g))/(\sigma_1(g) + \sigma_2(g)),$$

wherein g is the gene expression value; c is the class distinction, $\mu_1(g)$ is the mean of the expression levels for g for the first class; $\mu_2(g)$ is the mean of the expression levels for g for the second class; $\sigma_1(g)$ is the standard deviation for g the first class; and $\sigma_2(g)$ is the standard deviation for the second class.

57. In a computer system, a method for constructing a model for classifying at least one sample to be tested having a gene expression product, comprising:

a) receiving a vector for gene expression values of two or more samples belonging to more than one class, the vector being a series of gene expression values for the samples;

b) determining genes that are relevant for classification of a sample to be tested; and

c) using a weighted voting routine, constructing the model for classifying the samples using at least a portion of the genes determined in step B).

58. The method of Claim 57, wherein the step of determining employs a signal to noise routine, a Pearson correlation routine, or a Euclidean distance routine to determine the relevant genes.

59. The method of Claim 58, wherein the signal to noise routine is:

$$P(g,c)= (\mu_1(g)-\mu_2(g))/(\sigma_1(g) + \sigma_2(g)),$$

wherein g is the gene expression value; c is the class distinction, $\mu_1(g)$ is the mean of the expression levels for g for a first class; $\mu_2(g)$ is the mean of the expression levels for g for a second class; $\sigma_1(g)$ is the standard deviation for g the first class; and $\sigma_2(g)$ is the standard deviation for the second class.

60. The method of Claim 59, wherein the a weighted voting routine employs:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a negative vote for the class.

61. The method of Claim 60, wherein the vote for the first class is determined by obtaining a sum of the absolute values of the positive votes for the first class, and the vote for the second class is determined by obtaining a sum of the absolute values of the negative votes for the second class.

62. The method of Claim 61, further comprising performing cross-validation of the model.

63. The method of Claim 62, wherein performing cross-validation of the model comprises:

a) eliminating a sample used to build the model;

b) using a weighted voting routine, building a cross-validation model for classifying without the eliminated sample;

c) using the cross-validation model, classifying the eliminated sample including comparing the gene expression values of the eliminated sample to level of gene expression of the cross-validation model; and

d) determining a prediction strength of the class for the eliminated sample based on the cross-validation model classification of the eliminated sample.

64. The method of Claim 63, wherein the prediction strength is:

$$PS= (V_{win}-V_{lose})/V_{win}+V_{lose})$$

wherein V_{win} is the number of votes for the class to which the sample belongs, and V_{lose} the number of votes for the class to which the sample does not belong.

65. The method of Claim 57, further comprising filtering out any gene expression values in the sample that exhibit an insignificant change.

66. The method of Claim 57, further comprising normalizing the gene expression value of the vectors.

67. A method for ascertaining at least one previously unknown class into which at least one sample to be tested is classified, wherein the sample is obtained from an individual, comprising:

a) obtaining gene expression levels for a plurality of genes from two or more samples;

b) forming respective vectors of the samples, each vector being a series of gene expression values indicative of gene expression levels for the genes in a corresponding sample; and

c) using a clustering routine, grouping vectors of the samples such that vectors indicative of similar gene expression levels are clustered together to form working clusters, said working clusters defining at least one previously unknown class.

68. The method of Claim 67, wherein at least one previously unknown class is an unknown disease class; further comprising:

a) using a model built with a weighted voting scheme, classifying at least one sample by comparing gene expression levels of the sample to the model, such that a model classification results; and

b) using the model classification, validating at least one previously unknown disease class.

69. The method of Claim 67, wherein the clustering routine comprises a self organizing map.

70. The method of Claim 69, wherein the self organizing map is formed of a plurality of Nodes, N, and clusters the vectors according to a competitive learning routine.

71. The method of Claim 70, wherein the competitive learning routine is:

$$f_{i+1}(N) = f_i(N) + \tau(d(N,N_p), i)\,(P - f_i(N))$$

wherein i = number of iterations, N= the node of the self organizing map, τ = learning rate, P = the subject working vector, d = distance, N_p = node that is mapped nearest to P, and $f_i(N)$ is the position of N at i.

72. The method of Claim 71, wherein the routine for building a model with a weighted voting scheme is:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a negative vote for the class.

73. The method of Claim 69, further comprising filtering out any vectors that exhibit an insignificant change in the gene expression value, such that working vectors remain.

74. The method of Claim 69, further comprising normalizing the gene expression value of the working vectors.

75. The method of Claim 69, further comprising rescaling the gene expression values to account for variations across multiple conditions or experiments.

76. The method of Claim 69, further comprising providing an output indicating the formed working clusters.

77. The method of Claim 69, further comprising subjecting the sample to a condition or agent.

78. A method for ascertaining at least one previously unknown disease class into which at least one sample to be tested is classified, wherein the sample is obtained from an individual, comprising:

a) obtaining gene expression levels for a plurality of genes from two or more samples;

b) forming respective vectors of the samples, each vector being a series of gene expression values indicative of gene expression levels for the genes in a corresponding sample; and

c) using a clustering routine, grouping vectors of the samples such that vectors indicative of similar gene expression levels are clustered together to form working clusters, said working clusters defining at least one previously unknown disease class.

79. The method of Claim 78, further comprising:

a) using a computer model built with a weighted voting scheme, classifying at least one sample by comparing gene expression levels of the sample to the model, such that a model classification results; and

b) using the model classification, validating at least one previously unknown disease class.

80. The method of Claim 78, wherein the unknown disease class is a proliferative disease class.

81. The method of Claim 80, wherein the proliferative disease is cancer.

82. The method of Claim 81, wherein cancer is leukemia.

83. A computer apparatus for classifying a sample into a class, wherein the sample is obtained from an individual, wherein the apparatus comprises:

a) a source of gene expression values of the sample;

b) a processor routine executed by a digital processor, coupled to receive the gene expression values from the source, the processor routine determining classification of the sample by comparing the gene expression values of the sample to a model built with a weighted voting scheme; and

c) an output assembly, coupled to the digital processor, for providing an indication of the classification of the sample.

84. The computer apparatus of Claim 83, wherein the model is built according to:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a negative vote for the class.

85. The computer apparatus of Claim 84, wherein the vote for the first class is determined by obtaining a sum of the absolute values of the positive votes for the first class, and the vote for the second class is determined by obtaining a sum of the absolute values of the negative votes for the second class.

86. The computer apparatus of Claim 84, wherein the output assembly comprises a display of the classification.

87. A computer apparatus for constructing a model for classifying at least one sample to be tested having a gene expression product, wherein the apparatus comprises:

a) a source of vectors for gene expression values from two or more samples belonging to two or more classes, the vector being a series of gene expression values for the samples;

b) a processor routine executed by a digital processor, coupled to receive the gene expression values of the vectors from the source, the processor routine determining relevant genes for classifying the sample, and constructing the model with a portion of the relevant genes by utilizing a weighted voting scheme.

88. The computer apparatus of Claim 87, further comprising an output assembly, coupled to the digital processor, for providing the model.

89. The computer apparatus of Claim 87, wherein a weighted voting routine employs:

$$V_g=a_g(x_g-b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g=\mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a negative vote for the class.

90. The computer apparatus of Claim 89, wherein the vote for the first class is determined by obtaining a sum of the absolute values of the positive votes for the first class, and the vote for the second class is determined by obtaining a sum of the absolute values of the negative votes for the second class.

91. The computer apparatus of Claim 89, wherein the relevant genes are determined by a signal to noise routine, a Pearson correlation routine, or a Euclidean distance routine.

92. The computer apparatus of Claim 91, wherein the signal to noise routine is:

$$P(g,c)= (\mu_1(g)-\mu_2(g))/(\sigma_1(g) + \sigma_2(g)),$$

wherein g is the gene expression value; c is the class distinction; $\mu_1(g)$ is the mean of the expression levels for g for the first class; $\mu_2(g)$ is the mean of the expression levels for g for the second class; $\sigma_1(g)$ is the standard deviation for g the first class; and $\sigma_2(g)$ is the standard deviation for the second class.

93. The computer apparatus of Claim 87, further comprising a filter, coupled between the source and the processor routine, for filtering out any of the gene expression values in a sample that exhibit an insignificant change.

94. The computer apparatus of Claim 87, further comprising a normalizer, coupled to the filter, for normalizing the gene expression values.

95. The computer apparatus of Claim 87, wherein the output assembly comprises a display of the model.

96. The computer apparatus of Claim 95, wherein the output assembly comprises a graphical representation.

97. The computer apparatus of Claim 96, wherein the graphical representation is color coordinated.

98. The computer apparatus of Claim 97, wherein the color coordination comprises shades of contiguous colors.

99. A computer apparatus for ascertaining at least one previously unknown class into which at least one sample to be tested is classified, wherein the sample is obtained from an individual, comprising:

a) a source of gene expression values for a plurality of genes from two or more samples, for each sample, a series of gene expression values for the genes in the sample forms a vector; and

b) a processor routine, executed by a digital processor, coupled to receive the gene expression values from the source, the processor routine clustering vectors of the samples such that vectors indicative of similar gene expression levels are clustered together to form working clusters, said working clusters defining at least one previously unknown class.

100. The computer apparatus of Claim 99, wherein the processor routine employs a model built with a weighted voting scheme to classify the sample by comparing gene expression levels of the sample to the model such that a model classification results and, using the model classification, validating the at least one previously unknown class.

101. The computer apparatus of Claim 99, wherein the vectors are clustered with a self organizing map.

102. The computer apparatus of Claim 101, wherein the self organizing map is formed of a plurality of Nodes, N, and clusters the vectors according to a competitive learning routine.

103. The computer apparatus of Claim 102, wherein the competitive learning routine is:

$$f_{i+1}(N) = f_i(N) + \tau(d(N,N_p), i)\,(P - f_i(N))$$

wherein i = number of iterations, N= the node of the self organizing map, τ = learning rate, P = the subject working vector, d = distance, N_p = node that is mapped nearest to P, and $f_i(N)$ is the position of N at i.

104. The computer apparatus of Claim 100, wherein the weighted voting scheme is:

$$V_g = a_g(x_g - b_g),$$

wherein V_g is the weighted vote of the gene, g; a_g is the correlation between gene expression values and class distinction; $b_g = \mu_1(g)+\mu_2(g))/2$ which is the average of the mean $\log_{10}$ expression value in a first class and a second class; x_g is the $\log_{10}$ gene expression value in the sample to be tested; and wherein a positive V value indicates a vote for the first class, and a negative V value indicates a negative vote for the class.

105. The computer apparatus of Claim 99, further comprising a filter, coupled between the source and the processor routine, for filtering out any vectors

that exhibit an insignificant change in the gene expression value, such that working vectors remain.

106. The computer apparatus of Claim 99, further comprising a normalizer, coupled to the filter, for normalizing the gene expression value of the working vectors.

107. A machine readable computer assembly for classifying a sample into a class, wherein the sample is obtained from an individual, wherein the computer assembly comprises:

a) a source of gene expression values of the sample;

b) a processor routine executed by a digital processor, coupled to receive the gene expression values from the source, the processor routine determining classification of the sample by comparing the gene expression values of the sample to a model built with a weighted voting scheme; and

c) an output assembly, coupled to the digital processor, for providing an indication of the classification of the sample.

108. A machine readable computer assembly for constructing a model for classifying at least one sample to be tested having a gene expression product, wherein the computer assembly comprises:

a) a source of vectors for gene expression values from two or more samples belonging to two or more classes, the vector being a series of gene expression values for the samples;

b) a processor routine executed by a digital processor, coupled to receive the gene expression values of the vectors from the source, the processor routine determining relevant genes for classifying the

sample, and constructing the model with a portion of the relevant genes by utilizing a weighted voting scheme.

109. A method of determining a treatment plan for an individual having a disease, comprising:

a) obtaining a sample from the individual;

b) assessing the sample for the level of gene expression for at least one gene;

c) using a computer model built with a weighted voting scheme, classifying the sample into a disease class, as a function of relative gene expression level of the sample with respect to that of the model; and

d) using the disease class, determining a treatment plan.

110. The method of Claim 109, wherein the disease is cancer.

111. A method of diagnosing or aiding in the diagnosis of an individual, wherein a sample from the individual is obtained, comprising:

a) assessing the sample for the level of gene expression for at least one gene; and

b) using a computer model built with a weighted voting scheme, classifying the sample into a class of the disease including evaluating the gene expression level of the sample with respect to gene expression level of the model; and

c) diagnosing or aiding in the diagnosis of the individual.

112. A method for determining a drug target of a condition or disease of interest (i.e, genes that are relevant/important for a particular class), wherein a sample is obtained from an individual, comprising:

a) assessing the sample for the level of gene expression for at least one gene; and

b) using a neighborhood analysis routine, determining genes that are relevant for classification of the sample, to thereby ascertain a drug target.

113. The method of Claim 112, using a weighted voting routine, building or constructing a model for classifying the sample using at least a portion of the genes determined in step B).

114. A method of determining the efficacy of a drug designed to treat a disease class, comprising:

a) obtaining a sample from an individual having the disease class;

b) subjecting the sample to the drug;

c) assessing the drug exposed sample for the level of gene expression for at least one gene; and

d) using a computer model built with a weighted voting scheme, classifying the drug exposed sample into a class of the disease as a function of relative gene expression level of the sample with respect to that of the model.

115. A method of determining the efficacy of a drug designed to treat a disease class, wherein an individual has been subjected to the drug, comprising:

a) obtaining a sample from the individual subjected to the drug;

b) assessing the sample for the level of gene expression for at least one gene; and

c) using a model built with a weighted voting scheme, classifying the sample into a class of the disease including evaluating the gene

expression level of the sample as compared to gene expression level of the model.

116. A method of determining whether an individual belongs to a phenotypic class, comprising:

a) obtaining a sample from the individual;

b) assessing the sample for the level of gene expression for at least one gene; and

c) using a model built with a weighted voting scheme, classifying the sample into a class of the disease including evaluating the gene expression level of the sample as compared to gene expression level of the model.

117. The method of Claim 116, wherein the phenotypic class is selected from the group consisting of: intelligence, response to a treatment, length of life, likelihood of viral infection and obesity.

118. The method of Claim 18, wherein the cancer disease class is glioblastoma or medulloblastoma.

119. The method of Claim 18, wherein the cancer disease class is folicular lymphoma or diffuse large B cell lymphoma.

120. The method of Claim 37, wherein the cancer is a brain tumor.

121. The method of Claim 120, wherein the brain tumor is medulloblastoma or glioblastoma.

122. The method of Claim 37, wherein the cancer is Non-Hodgkin's lymphoma.

123. The method of Claim 122, wherein the lymphoma is folicular lymphoma or diffuse large B cell lymphoma.